Geophysical Prospecting **41,** 565–578, 1993

RESIDUAL STATICS ESTIMATION: SCALING TEMPERATURE SCHEDULES USING SIMULATED ANNEALING[1]

E. NØRMARK[2] and K. MOSEGAARD[3]

ABSTRACT

NØRMARK, E. and MOSEGAARD, K. 1993. Residual statics estimation: scaling temperature schedules using simulated annealing. *Geophysical Prospecting* **41,** 565–578.

Linearized residual statics estimation will often fail when large static corrections are needed. Cycle skipping may easily occur and the consequence may be that the solution is trapped in a local maximum of the stack-power function. In order to find the global solution, Monte Carlo optimization in terms of simulated annealing has been applied in the stack-power maximization technique. However, a major problem when using simulated annealing is to determine a critical parameter known as the temperature.

An efficient solution to this difficulty was provided by Nulton and Salamon (1988) and Andresen *et al.* (1988), who used statistical information about the problem, acquired during the optimization itself, to compute near optimal annealing schedules.

Although theoretically solved, the problem of finding the Nulton–Salamon temperature schedule often referred to as the schedule at constant thermodynamic speed, may itself be computationally heavy. Many extra iterations are needed to establish the schedule.

For an important geophysical inverse problem, the residual statics problem of reflection seismology, we suggest a strategy to avoid the many extra iterations. Based on an analysis of a few residual statics problems we compute approximations to Nulton–Salamon schedules for almost arbitrary residual statics problems. The performance of the approximated schedules is evaluated on synthetic and real data.

INTRODUCTION

When making reflection seismic surveys on land, different thicknesses and velocities of the surface layers will induce different delays on the seismic recordings, which

[1] Received April 1992, revision accepted November 1992.

[2] Department of Earth Sciences, Geophysical Laboratory, University of Aarhus, Finlandsgade 8, 8200 Aarhus N, Denmark.

[3] Geophysical Institute, University of Copenhagen, Haraldsgade 6, 2200 Copenhagen N, Denmark.

may generate false structures and reduce the quality of the common midpoint (CMP) stack. To compensate for these timing errors, static timeshifts of the observations are made. If the *a priori* information in terms of the field statics is insufficient, residual statics are estimated by automatic static correction procedures.

Usually, residual statics estimation is approached by a technique described by Taner, Koehler and Alhilali (1974), sometimes referred to as the traveltime picking method. This procedure is based on estimating timeshifts for all the individual traces in the CMP gathers and later resolving the time lags into surface-consistent source and receiver statics by least-squares fitting. Ronen and Claerbout (1985) have suggested an alternative statics estimation technique maximizing the stack-power S(**x**), where $\mathbf{x} = (x_1, x_2, \ldots, x_M)$ is the vector of the M source and receiver static parameters. In this method surface-consistent statics are estimated directly by evaluating one static parameter at a time.

The change in stack-power $\Delta S(\delta x_j)$ for a perturbation δx_j of the jth static parameter is

$$\Delta S(\delta x_j) = 2 \sum_{c \in \mathbf{C}_j} \phi^{pf}_{c(\delta x_j)}, \tag{1}$$

where $\phi^{pf}_{c(\delta x_j)}$ is the cross-correlation between the trace f affected by the jth static parameter and the partial stack p of the cth CMP gather. The partial stack is the stack of all traces in the CMP gather except the trace being studied. $\mathbf{C}_j$ is the subset of CMP gathers affected by the static parameter x_j.

For further details about the traveltime picking method and the stack-power maximization technique see Nørmark (1993).

Normally, the solution is approached by local optimization in both residual statics estimation techniques. The permitted timeshifts are evaluated and the static displacement giving the best correlation between the actual trace and a constructed reference trace (normally the partial stack) is applied to the data. However, a major problem is that cycle skipping often occurs when large residual statics, compared to the dominating period in the data, are estimated. The consequence may easily be that the solution is trapped in a local maximum of the objective function. This presents a highly non-linear inverse problem, which requires a global optimization technique to solve it.

Global Optimization by Simulated Annealing

Residual statics estimation considered as a global optimization problem was first treated by Rothman (1985). The optimization problem was approached by a Monte Carlo optimization technique, rooted in statistical mechanics, called simulated annealing. Simulated annealing is a numerical technique that resembles chemical annealing, in the way crystals are grown from a melt. This process is characterized by the fact that if the melt is carefully cooled in a critical temperature interval, a regular crystal with a minimum of energy is formed, whereas if the temperature is lowered too quickly, glass may be the result. A similarity between a thermodynamic description of such processes and the task of combinatorial optimization of a non-

linear function of high dimension was discovered by Kirkpatrick, Gelatt and Vecchi (1983). Based on this analogy they introduced simulated annealing as a tool to locate near-optimum solutions to global optimization problems.

Minimization of a parameter-dependent objective function $E(\mathbf{x})$ by simulated annealing is accomplished by making random perturbations of the parameters (corresponding to the state space coordinates in the physical problem) according to the Metropolis algorithm. The Metropolis algorithm states that, in each iteration, a random perturbation of a parameter, causing a change in the objective function (the energy) given by $\Delta E(\delta x_j)$, is

(1) accepted if $\Delta E(\delta x_j) \leq 0$,
(2) rejected with the probability $P(\delta x_j) = \exp(-\Delta E(\delta x_j)/T)$ if $\Delta E(\delta x_j) > 0$

(Metropolis *et al.* 1953). The parameter T corresponds to the temperature in thermodynamics, and for convenience is also named so in this context. When T is infinitely large, pure random perturbations of the parameters are made, whereas at $T = 0$ only perturbations decreasing the energy are accepted. Simulated annealing is initiated at high temperature. By slowly lowering the temperature and perturbing the parameters according to the Metropolis algorithm, the global minimum is reached with a high probability. Experimental evidences show the simulated annealing is much more efficient in locating the global minimum than the crude Monte Carlo optimization see e.g. Jakobsen, Mosegaard and Pedersen (1987).

Traditionally, only minimization problems are considered in simulated annealing. To keep the same convention in the residual statics estimation problem, the negative stack-power is minimized, which is equivalent to maximizing the stack-power. Thus $E(\mathbf{x}) = -S(\mathbf{x})$.

In the stack-power maximization problem, Rothman (1986) applies a modified version of the Metropolis algorithm, known as the heat bath method, which is claimed to be more efficient for problems where the energy evaluations are computationally inexpensive. In this method, the random trials are chosen according to the marginal Gibbs–Boltzmann probability distribution

$$P(\delta x_j) = \frac{\exp(\Delta E(\delta x_j)/T)}{\sum_{h=1}^{N} \exp(\Delta E(\delta x_h)/T)}, \tag{2}$$

where N is the number of possible static corrections for each parameter. $\Delta E(\delta x_j)$ (or $-\Delta S(\delta x_j)$) is evaluated using (1). By repeating the random perturbations according to the transition probabilities above (at constant temperature) and taking all static parameters once in each iteration, Rothman (1986) shows that eventually this will lead to a Boltzmann distribution of the available states, just as a repetition of the Metropolis algorithm will do.

The temperature schedule chosen is crucial for the performance of the optimization algorithm, in the sense that lowering the temperature too slowly will be a waste of computer time and cooling the system too fast will most likely result in a solution trapped in a local minimum. The temperature schedule suggested by

Rothman (1986) takes the form

$$T = \begin{cases} a^k T_0, & a^k T_0 > T_{\min}, \\ T_{\min}, & \text{otherwise}, \end{cases}$$

where k is the iteration number and α is a constant (set at 0.99). Firstly, a few iterations (controlled by T_0) are made with exponential cooling and afterwards a constant temperature $T_{\min}$ is applied at which essentially all iterations are made. The establishment of the above schedule is based on experiments.

Nulton and Salamon (1988) and Andresen *et al.* (1988) described a method by which near-optimum annealing temperature schedules could be produced. Their idea was to extend the analogy between the Monte Carlo optimization algorithms and statistical mechanics. Nulton and Salamon (1988) defined the heat capacity and the relaxation time for a problem and used principles from finite-time thermodynamics to design annealing schedules with minimum 'entropy' production i.e. annealing schedules at constant thermodynamic speed. Andresen *et al.* (1988) provided a method for the numerical estimation of constant speed schedules. In their method, attempted energy transitions are monitored during annealing, and from this information they estimate heat capacity and relaxation time for the considered problem. (See Appendix A). Constant speed schedules can then be calculated.

The Nulton–Salamon method has proved its efficiency in many cases (see e.g. Mosegaard and Vestergaard (1991)), but there are two important, practical problems in using this method.

Firstly, the Andresen *et al.* (1988) procedure is rather difficult to implement and requires a great deal of experience to use it. Secondly, the amount of extra iterations needed to give useful information about the heat capacity and relaxation time of a problem, following Andresen *et al.* (1988), may be so large, that the expected gain in computational efficiency is significantly reduced.

In order to overcome these problems in large residual statics estimation, we shall in the following suggest a strategy in which we compute a Nulton–Salamon schedule for a single (or a few) representative residual statics problems and apply a simple scaling procedure to approximate Nulton–Salamon schedules for other residual statics problems.

The Objective Function

Let us first consider the trivial problem of transferring temperature schedules from one problem to another, when the optimization problems are defined in the same parameter space and when the energies are linearly related. Then for two problems (1 and 2) the energies are related by

$$E_1(\mathbf{x}) = \alpha E_2(\mathbf{x}) + \beta, \tag{3}$$

where α and β are constants. Considering the same transition of state in both problems we have

$$\Delta E_1 = \alpha \, \Delta E_2 .$$

If we apply simulated annealing in the two problems with temperature schedules $T_1(t)$ and $T_2(t)$ respectively, so that

$$T_1(t) = \alpha T_2(t),$$

it is obvious that the optimizations become identical, because the parameter perturbations follow the same probability distributions in the two cases (see (2)).

In order to study whether a linear transformation of the temperature schedules is approximately valid for a broader class of residual statics estimation problems, not necessarily defined in parameter spaces with the same dimension and not necessarily satisfying (3), the following considerations are made:

Let the CMP gather c consist of M traces identified by index i. The individual traces are assumed to carry the signals s_{cit} and the noise n_{cit}, and are displaced by varying timeshifts $k(ci)$. t indicates the sample number. The CMP stack g_{ct} can be expressed as

$$g_{ct} = g_{ct}^{\text{signal}} + g_{ct}^{\text{noise}},$$

where

$$g_{ct}^{\text{signal}} = \sum_{i=1}^{M} \delta_{tk(ci)} * S_{\text{cit}} \qquad \text{and} \qquad g_{ct}^{\text{noise}} = \sum_{i=1}^{M} \delta_{tk(ci)} * n_{\text{cit}}.$$

$\delta_{tk(ci)}$ is Kronecker's delta given by

$$\delta_{tk(ci)} = \begin{cases} 1, & t = k(ci), \\ 0, & t \neq k(ci), \end{cases}$$

describing the static displacements.

If it is assumed that the signals are uncorrelated with the noise, the power of the CMP stack $S_c = \sum_t (g_{ct})^2$ can be approximated by

$$S_c \approx S_c^{\text{signal}} + S_c^{\text{noise}}.$$

Let us subdivide the stack-power range into intervals. The ith interval contains stack-powers between S_i and $S_i + \delta S$, where δS is the interval width. During a run, we can now form a matrix $\mathbf{A} = \{a_{ij}\}$ where a_{ij} is the number of transitions that could have taken place from the ith stack-power level to the jth stack-power level. In every iteration, we start at a certain stack-power level i and are allowed to perform a transition to any state that can be reached by only changing one static parameter. In such an iteration, all the reachable statics contribute to $\mathbf{A}$. The ith row in the matrix $\mathbf{A}$ will, after normalization, contain the distribution of potential stack-power transitions starting in stack-power level i. Let $d(S_i)$ be the standard deviation of this distribution.

By assuming that the stack-power perturbations have a Gaussian distribution, we obtain

$$d^2 \approx d_{\text{signal}}^2 + d_{\text{noise}}^2,$$

where d_{signal} and d_{noise} are the standard deviations for the noise and the signals respectively.

As will be seen the magnitude of the potential stack-power perturbations is proportional to the average stack-power associated with the static parameters. That is

$$d(\langle S_c \rangle) = aC\langle S_c \rangle + b,$$

where $\langle S_c \rangle$ is the average stack-power per CMP, and C is the average number of members in $\mathbf{C}_j$, the subset of the CMP gather over which the stack-power changes are evaluated (see (1)). a and b are constants.

In order to verify this relationship, and in order to estimate a and b, two optimization experiments were made, one on noise-free data and one on pure noise. 6-fold synthetic seismic data are used on which simulated annealing is applied. In Fig. 1 the standard deviations d of the potential stack-power perturbations are shown as a function of the stack-power for both experiments. Both d and S are normalized with the maximum stack-power being observed.

It can be seen that a linear relationship between stack-power and the standard deviation of the potential stack-power perturbations is a good approximation, although slight deviation from this relationship is observed at low stack-power, especially for the data containing pure noise. In both cases the intersections of the linear fits with the horizontal axes are approximately equal to the minimum stack-power being observed. Thus

$$d_{\text{signal}} \approx a_{\text{signal}} C(\langle S_c \rangle^{\text{signal}} - \langle S_c \rangle^{\text{signal}}_{\text{min}})$$

and

$$d_{\text{noise}} \approx a_{\text{noise}} C(\langle S_c \rangle^{\text{noise}} - \langle S_c \rangle^{\text{noise}}_{\text{min}}).$$

In the experiments it was found that $a_{\text{signal}} \approx a_{\text{noise}}$.

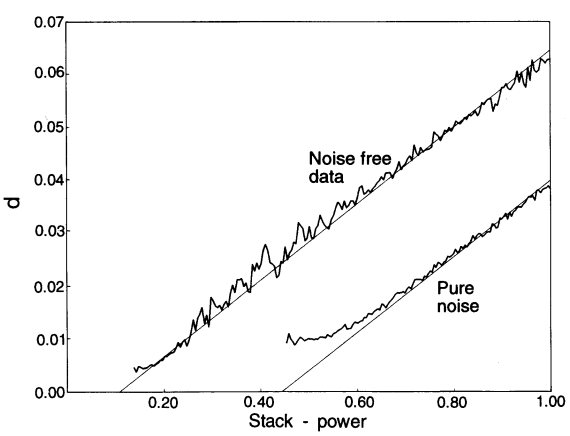

FIG. 1. Standard deviation of the potential stack-power perturbations d for experiments on noise-free data and on pure noise. The stack-power perturbations are mapped as a function of the stack-power itself. Both the stack-power and the standard deviations are normalized by the maximum stack-power. The linear fits are indicated by thin lines.

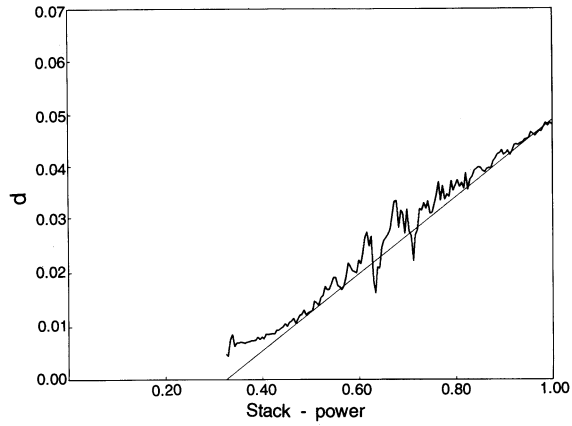

FIG. 2. The standard deviation d of the potential stack-power perturbations for data with a signal-to-noise ratio of 1 on the unstacked data. As in Fig. 1 the stack-power perturbations are mapped as function of the stack-power. The straight line indicates the predicted stack-power perturbation.

If it is assumed that signal and noise are uncorrelated, we have

$$S_c^{\text{signal}} \approx (1 + P^{-2})^{-1} S_c \quad \text{and} \quad S_c^{\text{noise}} \approx (1 + P^2)^{-1} S_c,$$

where $P = (S_c^{\text{signal}})^{1/2}/(S_c^{\text{noise}})^{1/2}$ is the signal-to-noise ratio of the CMP stack.

In the general case of noise-contaminated data, an estimate of d is given by

$$\begin{aligned} d^2 &\approx (a_{\text{signal}} C(\langle S_c \rangle^{\text{signal}} - \langle S_c \rangle_{\min}^{\text{signal}}))^2 + (a_{\text{noise}} C(\langle S_c \rangle^{\text{noise}} - \langle S_c \rangle_{\min}^{\text{noise}}))^2 \\ &\approx ((1 + P^{-2})^{-1} a_{\text{signal}} + (1 + P^2)^{-1} a_{\text{noise}})^2 (C(\langle S_c \rangle - \langle S_c \rangle_{\min}))^2. \end{aligned}$$

Since a is similar for both the signal and the noise experiments, P vanishes and d is approximated by

$$d = aC(\langle S_c \rangle - \langle S_c \rangle_{\min}). \tag{4}$$

In order to confirm (4), an experiment on noise-contaminated data is made, similar to those experiments made on noise-free data and on pure noise. The signal-to-noise ratio is 1 on the unstacked data within the cross-correlation window. In Fig. 2 the observed standard deviations of the potential stack-power perturbations are shown, together with the standard deviations predicted from (4). It can be seen that the estimated and the observed deviations are in good agreement with each other.

NORMALIZING THE TEMPERATURE SCHEDULES

From (2) it seems promising to obtain approximate Nulton–Salamon schedules by normalizing the temperatures by the magnitude of the expected energy perturbations, which we evaluated as the standard deviation of the potential stack-power

changes d. The aim of this section is to estimate the temperature schedules for a small, but hopefully reasonably representative, group of residual statics estimation problems, and normalize them to form schedules for other residual static problems.

According to (4), which has been verified experimentally, d varies approximately linearly with the stack-power. Let us consider two different static problems (1 and 2). During annealing, the Nulton–Salmon schedule $T_1(t)$ for problem 1 will increase the stack-power from $S_{1\,\min}$ to $S_{1\,\max}$ in a given number of iterations. The schedule $T_2(t)$ for problem 2 should increase the stack-power for that problem from $S_{2\,\min}$ to $S_{2\,\max}$ over the same number of iterations. Let us assume further that the stack-power functions (in the parameter space) for the two problems are different realizations of the same stochastic process, except for a linear transformation such as (3). In any iteration, we therefore require that the typical Gibbs–Boltzmann probabilities of (2) are approximately the same for both problems. This will be the case if the relationship

$$\frac{d_1}{T_1} = \frac{d_2}{T_2}$$

is satisfied. Here, d_1 and d_2 are the standard deviations of the potential stack-power perturbations for problems 1 and 2 respectively. It is readily seen that the above realization can be satisfied by the scaling

$$\frac{d_{1\,\max}}{T_1} = \frac{d_{2\,\max}}{T_2},$$

where $d_{1\,\max}$ and $d_{2\,\max}$ are d_1 and d_2 at the maximum stack-powers $S_{1\,\max}$ and $S_{2\,\max}$ respectively.

Normally, $d_{\max}$ is inaccessible, but according to (4) one may normalize by the stack-power instead. We use $C(\langle S_c\rangle_{\max} - \langle S_c\rangle_{\min})$ as a scaling factor. a vanished as it was found to be almost problem independent. In most residual static problems $\langle S_c\rangle_{\max}$ and $\langle S_c\rangle_{\min}$ are unknown, but a reasonable guess is usually possible.

In order to estimate the validity of the scaling suggested above, a number of schedules are estimated on synthetic examples with different numbers of parameters, varying signal-to-noise ratios, different stack-folds and different maximum static corrections. A schedule based on a real seismic data set is also included. Seismic data with between 6- and 12-fold coverages are employed. All examples have a fairly high signal-to-noise ratio. The input data are normalized in such a way that the mean power of the unstacked traces is equal to all examples. In all experiments a number of static parameters are fixed to the assumed solution at the end of the profile before initiating the optimization. This makes it easier for the algorithm to start aligning the seismic events. In all optimization problems the static parameters are taken in random order. The resulting temperature schedules are shown in Fig. 3a. No normalization has been applied. When normalizing by $C(\langle S_c\rangle_{\max} - \langle S_c\rangle_{\min})$ as in Fig. 3b, it is seen that the schedules have become much closer. It will now be investigated whether cooling curves normalized in this way can be used for the construction of the master schedule, from which approximate Nulton–Salamon schedules for other static problems can be derived.

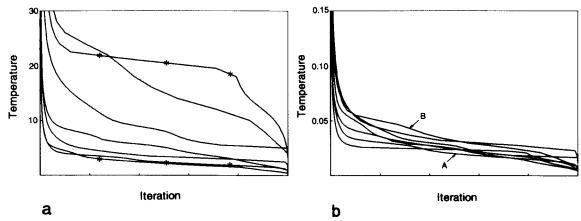

FIG. 3. (a) Temperature before normalization; (b) after normalization has been applied.

The data examples A and B, providing schedules with the lowest and highest normalized temperatures, are analysed. Example A is constructed from a synthetic seismic data set carrying a minimum phase signal and example B originates from real seismic data on which random statics are applied. In the latter case it was decided to use data where originally no significant statics problems existed. This makes an evaluation of the static solution easier. Figure 4a shows the initial states for the optimization and Fig. 4b shows the true solutions. In Fig. 4c, two examples of carrying out local optimizations are given. Almost all experiments with local optimization gave solutions trapped in local maxima of the stack-power function.

In order to demonstrate the possible consequence of using non-normalized temperature schedules estimated from another experiment, the schedules with highest and lowest temperatures (indicated by * in Fig. 3a) have been applied on data set A.

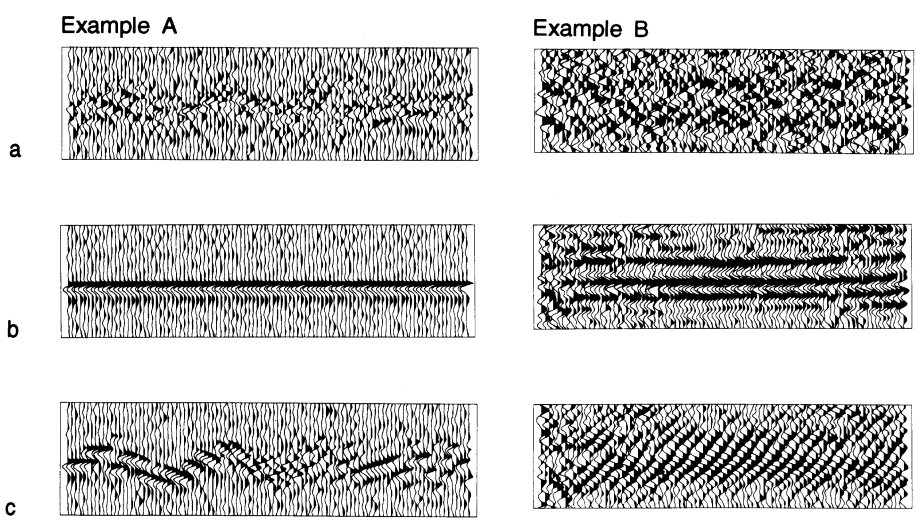

FIG. 4. Local optimization on two data sets. A is constructed from synthetic seismic data and B originates from a real seismic data set. (a) The initial state of the optimization; (b) the true solution; (c) examples of making local optimization by taking the static parameters in random order.

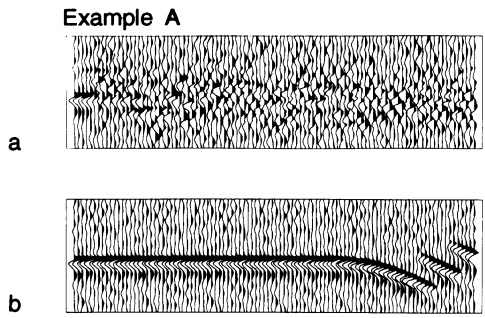

FIG. 5. The outcome of simulated annealing using (unnormalized) temperature schedules estimated from other experiments. (a) A schedule is used with temperatures that are generally too high. No order in the seismic traces has been achieved. (b) The temperatures are too low for the present data. The solution is trapped in a local maximum of the stack-power function.

In Fig. 5 the outcome of simulated annealing with 600 iterations is illustrated. (One iteration refers here to a perturbation of all static parameters). It is observed that when using a temperature schedule that is too high, no order in the seismograms has been detected (Fig. 5a). By using a schedule with temperatures that are too low, convergence is achieved after 530 iterations (Fig. 5b). Apparently, too rapid cooling has taken place through the most critical temperatures and the solution is trapped in a local optimum. This clearly demonstrates the need for normalizing the temperature schedules.

The normalized schedules in Fig. 3b still show a rather large scattering of the temperatures. In order to study the significance of these variations and thus the validity of a master temperature schedule based on an average of these cooling curves, the following experiments are made. First, simulated annealing on both data sets A and B is made using their own temperature schedules. 600 iteratons are made on 5 copies of the same data set. Representative solutions are shown in Fig. 6a. It is observed that the local maxima of the stack-power have now been avoided. By repeating these two experiments and changing their temperature schedules the importances of the differences between the schedules can be studied. The cooling curves are based on the normalized temperatures, which are scaled to match the actual data. The outcome of the optimizations is shown in Fig. 6b. Practically, the same solutions are obtained as when using their own temperature schedules. Only a minor effect of employing another (normalized) schedule than its own, is observed on the static corrections as well as on the stack-power function. For instance, on the synthetic data example, long periodic statics are still left by using a schedule estimated from another data set. However, one cannot expect to resolve long-periodic static by residual statics estimation (Wiggins, Larner and Wisecup 1976).

These experiments demonstrate that a temperature schedule for a given problem can successfully be estimated from another experiment, provided a proper normalization has been applied. They also show that slight variations on the schedules do not give rise to any significant variations in the appearance of the stacked data, at

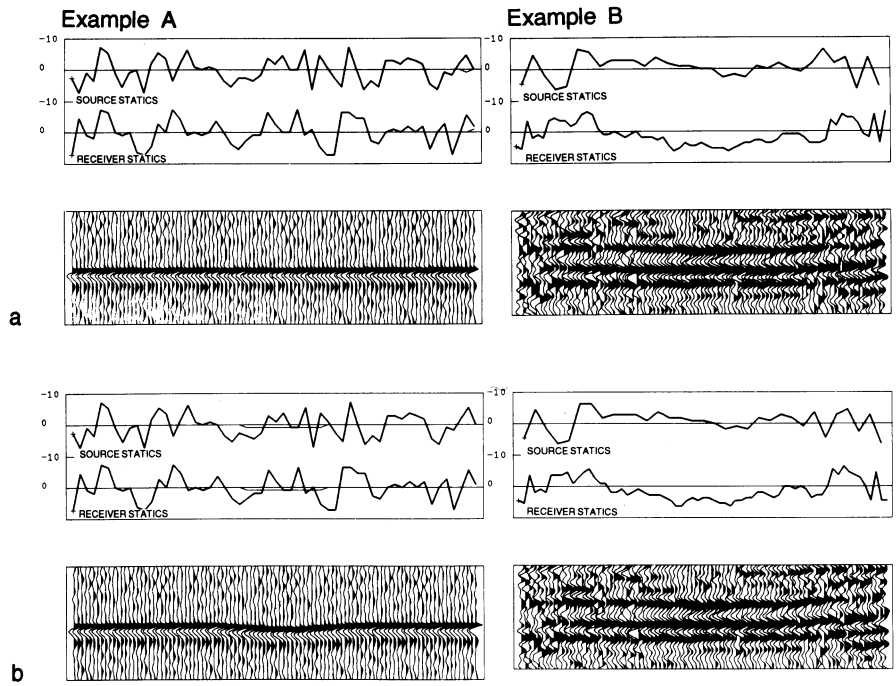

FIG. 6. (a) Simulated annealing using their own temperature schedules. (b) The result of changing their normalized temperature schedules. Before the schedules are used they are adapted to the actual data. The statics are given above the seismograms. The thin line in example A indicates the difference between the true and the estimated statics.

least not for the data examples studied in this context. Yet, the results of the optimization are not completely insensitive to such fluctuations. In order to construct a master schedule we suggest following an average schedule that is the mean value of the estimated master schedule.

DISCUSSION

Since the change in the energy for a static parameter x_j is only influenced by the CMP gathers in $\boldsymbol{C}_j$, we can expect that the temperatures are independent of the size of the seismic profile. However, when treating larger profiles local order may start to occur in different parts of the section, without global consistency. The consequence may be that more iterations are needed to find the global solution and that the temperature schedule must be modified accordingly. Yet, in this context only minor profiles, extending over a few spread lengths, have been treated. End effects causing decreasing coverage in the end of the profile and the resulting difficulties in estimating static parameters at such places have also been ignored.

Generally, it was found that normalized temperature schedules are fairly insensitive to variations in the input data. Experiments confirm that changing the stack-

fold does not have a significant effect on the normalized cooling schedules. However, only data of low coverages, i.e. up to 14-fold, have been examined. Using data of higher coverages will in itself make it more difficult to find a solution, because order in the seismograms is harder to achieve. Experiments also showed that changing the length of the cross-correlation window does not have any effect on the schedules. Modifying the maximum allowable static correction will also have no significant effect on the temperature schedules.

Our analysis of the temperature scaling problem shows that the temperatures generally decrease with increasing noise level due to the fact that the difference between $\langle S_c \rangle_{\max}$ and $\langle S_c \rangle_{\min}$ is small for noisy data. If statics estimation is made on data with a very high noise level, it may be hard to find any correlation at all between the seismic traces. In such cases it may be fruitful to scale the temperature schedules as if the data were pure noise.

Experimentally, it was discovered that when the source spacing deviates significantly from the receiver spacing the overall character of the temperature schedules may change. The numbers of traces in the source and the receiver gathers are then different, and consequently different magnitudes of the energy perturbations of the source and the receiver parameters can be expected, which may have a significant influence on the temperature schedule.

Conclusion

By studying the objective function of the residual statics estimation problem, it has been shown that the magnitude of the energy perturbations can be estimated and used to normalize the temperature schedules. Generally, it was found that normalized schedules are fairly consistent, and can, for a larger class of residual statics estimation problems, form a master temperature schedule from which approximate Nulton–Salamon schedules can be found. Our master schedules are based on a few static problems. A larger variety of problems should be included in order to determine a practically applicable master temperature schedule.

The present technique of calculating temperature schedules on the representative group of problems, studying the characteristics of the objective function and normalizing the temperatures in order to estimate a master schedule, could be used as the model for other non-linear optimization problems solved by simulated annealing.

Appendix A
Annealing with Constant Thermodynamic Speed

The philosophy of the present technique is to determine the temperature schedule keeping the same 'distance' to equilibrium mean energy $\langle E(T) \rangle_{\text{eq}}$ during the optimization. The 'distance' called the thermodynamic distance v, is defined as

$$v = \frac{\langle E(T) \rangle - \langle E(T) \rangle_{\text{eq}}}{\sigma(E_{\text{eq}}(T))},$$

where $\langle E(T) \rangle$ is the (non-equilibrium) mean energy of the states at temperature T and $\sigma(E(T)_{eq})$ is the standard deviation of energy fluctuations. (Nulton and Salamon 1988).

By adopting two other concepts from thermodynamics, the heat capacity $C(T)$ and the relaxation time $\varepsilon(T)$, and keeping v constant, the temperature schedule can be estimated as

$$\frac{\delta T}{\delta t} = -\frac{vT}{\varepsilon(T)\sqrt{C(T)}}, \tag{A1}$$

where t is the time measured in units of iterations. This first-order differential equation defines the temperature schedule for what is called annealing with constant thermodynamic speed. (See also Mosegaard and Vestergaard (1991)).

$C(T)$ and $\varepsilon(T)$ are determined by a method based on statistics of the energy transitions, (Andresen *et al.* 1988). The states of the system are lumped by the energy and the number of attempted moves from one energy interval i to another energy interval j is recorded in a matrix $\mathbf{Q} = \{Q_{ij}\}$, which is normalized to form a transition matrix. Using attempted moves means that both accepted perturbations and perturbations being rejected, according to the Metropolis algorithm, are employed. By transforming $\mathbf{Q}$ to the temperature-dependent transition probability probability matrix $\mathbf{Q}(T)$ and calculating its largest and second largest eigenvalues and the corresponding eigenvectors, an estimate of the density of states (from which the heat capacity can be defined) and the relaxation time can be obtained. See e.g. Mosegaard and Vestergaard (1991).

References

Andresen, B., Hoffmann, K.H., Mosegaard, K., Nulton, J., Pedersen, J.M. and Salmon, P. 1988. On lumped models for thermodynamical properties of simulated annealing problems. *Journal Physique (France)* **49**, 1485–1492.

Jakobsen, M.O., Mosegaard, K. and Pedersen, J.M. 1987. Model optimization in exploration geophysics 2: Global model optimization in reflection seismology simulated annealing. In: *Proceedings of the 5th International Mathematical Geophysics Seminar, Berlin*, 361–381. Vieweg & Sohn, Braunschweig.

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. 1983. Optimization by simulated annealing. *Science* **220**, 671–680.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. 1953. Equation of state calculations by fast computing machines. *Journal Chemical Physics* **21**, 1087–1092.

Mosegaard, K. and Vestergaard, P.D. 1991. A simulated annealing approach to seismic model optimization with sparse prior information. *Geophysical Prospecting* **39**, 599–611.

Nørmark, E. 1993. Residual statics estimation by stack-power maximization in the frequency domain. *Geophysical Prospecting* **41**, 551–563.

Nulton, J. and Salamon, P. 1988. Statistical mechanics of combinatorial optimization. *Physical Review* **A37**, 1351–1356.

Ronen, J. and Claerbout, J.F. 1985. Surface-consistent residual statics estimation by stack-power maximization. *Geophysics* **50**, 2759–2767.

ROTHMAN, D.H. 1985. Nonlinear inversion, statistical mechanics, and residual statics estimation. *Geophysics* **50**, 2784–2796.

ROTHMAN, D.H. 1986. Automatic estimation of large statics corrections. *Geophysics* **51**, 332–346.

TANER, M.T., KOEHLER, F. and ALHILALI, K.A. 1974. Estimation and correction of near-surface time anomalies. *Geophysics* **39**, 441–463.

WIGGINS, R.A., LARNER, K.L. and WISECUP, R.D. 1976. Residual statics analysis as a general linear inverse problem. *Geophysics* **42**, 922–938.